Supplementary Materials for
"Age-determined expression of priming protease TMPRSS2 and localization of SARS-CoV-2 in the epithelium"

Supplemental Acknowledgements

Vanderbilt COVID-19 Cohort Consortium (VC3):
Justin L. Balko, Suman Das, David Haas, Spyros A. Kalams, Jonathan A. Kropski, Christine Lovly, Simon A. Mallal, Elizabeth J. Phillips, Wei Zheng, Vanderbilt University Medical Center, Nashville Tennessee

Human Cell Atlas Lung Biological Network
The members of HCA Lung Biological Network are Nicholas E. Banovich (Translational Genomics Research Institute), Pascal Barbry (Universite Cote d'Azur), Alvis Brazma (EMBL-European Bioinformatics Institute), Tushar Desai (Stanford University), Thu Elizabeth Duong (University of California, San Diego), Oliver Eickelberg (University of Pittsburgh), Christine Falk (Hannover Medical School), Michael Farzan (Scripps Research Institute), Ian Glass (University of Washington), Muzlifah Haniffa (Wellcome Sanger Institute), Peter Horvath (University of Helsinki), Deborah Hung (Massachusetts General Hospital), Naftali Kaminski (Yale University), Mark Krasnow (Stanford University), Jonathan A. Kropski (Vanderbilt University), Malte Kuhnemund (Cartana AB), Robert Lafyatis (University of Pittsburgh), Haeock Lee (Catholic University of Korea), Sylvie Leroy (Universite Cote d'Azur), Sten Linnarson (Karolinska Institute), Joakim Lundeberg (KTH Royal Institute of Technology), Kerstin B. Meyer (Wellcome Sanger Institute), Alexander Misharin (Northwestern University), Martijn Nawijn (University of Groningen), Marko Z. Nikolic (University College London), Jose Ordovas-Montanes (Harvard Medical School), Dana Pe'er (Memorial Sloan Kettering Cancer Center), Joseph Powell (University of New South Wales), Stephen Quake (Stanford University), Jay Rajagopal (Massachusetts General Hospital), Purushothama Rao Tata (Duke University), Emma L. Rawlins (University of Cambridge), Aviv Regev (Massachusetts Institute of Technology), Paul A. Reyfman (Northwestern University), Mauricio Rojas (University of Pittsburgh), Orit Rosen (Broad Institute of Harvard and MIT) , Kourosh Saeb-Parsy (University of Cambridge), Christos Samakovlis (Stockholm University), Herbert Schiller (Helmholtz Zentrum Munchen), Joachim L. Schultze (University of Bonn), Max A. Seibold (National Jewish Health), Alex K. Shalek (Broad Institute of Harvard and MIT), Douglas Shepherd (Arizona State University), Jason Spence (University of Michigan), Avrum Spira (Boston University), Xin Sun (University of San Diego), Sarah Teichmann (University of Cambridge), Fabian Theis (Helmholtz Zentrum Munchen), Alexander Tsankov (Icahn School of Medicine at Mt Sinai), Maarten van den Berge (University of Groningen), Michael von Papen(Comma Soft AG, Bonn), Jeffrey Whitsett (Cincinnati Children's Hospital), Ramnik Xavier (Massachusetts General Hospital), Laure-Emmanuelle Zaragosi (Universite Cote d'Azur), and Kun Zhang (University of California, San Diego). Pascal Barbry, Alexander Misharin, Martijn Nawijn, and Jay Rajagopal serve as the coordinators.

Methods

Animal Care and Tissue Fixation: C57BL/6 mice were used for all experiments. Timed matings were performed as previously described(1) and mice were sacrificed at P0, P7, P14, or P64 for single cell RNA sequencing or lung block fixation. E18 lungs were isolated by removing pups from the mouse uterus and isolating lung tissue. E18 and P0 lungs were fixed in formalin. P7, P14, P64, 12 month, and 24 month old mouse lungs were inflation-fixed by gravity filling with 10% buffered formalin and paraffin embedded. This protocol was approved by the Institutional Animal Care and Use Committee of Vanderbilt University (Nashville, TN) and was in compliance with the Public Health Services policy on humane care and use of laboratory animals.

Single cell isolation and flow cytometry: At the indicated timepoints, lung lobes were harvested, minced, and incubated for 30 minutes at 37°C in dissociation media (RPMI-1640 with 0.7 mg/ml collagenase XI and 30 mg/ml type IV bovine pancreatic DNase). After incubation, lobes were passed through a wide bore pipet tip and filtered through a 40 μm filter. Single cell lung suspension was then counted, aliquoted, and blocked with CD-32 Fc block (BD cat #553142) for 20 minutes on ice. After 2% FBS staining buffer wash, cells were incubated with the conjugated primary antibodies anti-CD45 (BD cat # 559864) and anti-Ter119 (BD cat# 116211) as indicated below. In the same manner, fluorescence minus one controls were blocked and stained with the appropriate antibody controls. Cells from individual mice were then incubated with identifiable hashtags, resuspended in staining buffer, and treated with PI viability dye. CD45 negative, Ter119 negative, viable cells were collected by fluorescence associated cell sorting using a 70 μm nozzle on a 4-laser FACSAria III Cell Sorter. Both single and fluorescence-minus-one controls were used for compensation.

scRNA-seq library preparation and next-generation sequencing: ScRNA-seq libraries were generated using the 10X Chromium platform 5' library preparation kits (10X Genomics) following the manufacturer's recommendations and targeting 10,000 - 20,000 cells per sample. Next generation sequencing was performed on an Illumina Novaseq 6000. Reads with read quality less than 30 were filtered out and CellRanger Count v3.1 (10X Genomics) was used to align reads onto mm10 reference genome.

Ambient RNA filtering: Ambient background RNA were cleaned from the scRNA-seq data with "SoupX" (version 1.2.2, Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK)(2) ; in RStudio (version 1.2.5001, RStudio, Inc., Boston, Massachusetts, USA). Matrix files from CellRanger were read into the global environment and combined into a "SoupChannel." Data from the SoupChannel were passed to Seurat (version 3.1.4, New York Genome Center, New York City, New York, USA)(3, 4) for data normalization, identification of variable features, data scaling, principal component analysis, uniform manifold approximation and projection (UMAP) for dimensionality reduction, calculation of nearest neighbors, and cluster identification. The SoupX pipeline was used for each time point to determine which genes were most likely contributing to the ambient background RNA. We utilized the following genes to estimate the non-expressing cells, calculate the contamination fraction, and adjust the gene expression counts: *Dcn, Bgn, Aspn, Ecm2, Fos, Hbb-bs, Hbb-bt, Hba-a1, Hba-a2, Lyz1, Lyz2, Mgp, Postn, Scgb1a1*. For time points that were sorted to select for epithelial cells, the following genes were

added to the SoupX pipeline: *Sftpc, Hopx, Ager, Krt19, Cldn4, Foxj1, Krt5, Sfn, Pecam1*. New matrix files were created by SoupX and used for subsequent analyses in Scanpy.

Data integration and clustering. Data integration and clustering was performed using a standard Scanpy workflow (Scanpy v1.46)(5). Individual SoupX "cleaned" libraries were concatenated and analyzed jointly. After quality filtering (removal of cells with <500 or >5000 genes, or >10% mitochondrial gene expression), data were normalized, log-transformed, highly-variable genes were identified, and percent mitochondrial gene expression and read depth were regressed. Following data scaling, and principal components analysis, Leiden clustering was performed followed by visualization of canonical marker genes. Red blood cells and doublet clusters (containing non-physiologic marker combinations, i.e. *Epcam*+/*Pecam1*+) were filtered. Batch correction for dataset integration was performed using batch-balanced K-nearest neighbors(6). Principal components 1:30 were used for clustering and Uniform Manifold Approximation and Projection (UMAP) embedding(7). Leiden clustering (resolution 0.8) was then performed on the integrated dataset and cell-types assigned based on marker expression profiles.

Subjects and samples: Lung tissue from 20 healthy human subjects was obtained at the time of surgical biopsy or autopsy, with death occurring from non-respiratory causes. Subjects were classified as infants (between birth and age 2 years), children (3-17 years), and adult (54-69 years old). Adult tissue included male and female donors, and all were lifetime nonsmokers. SARS-CoV-2 lung tissue was obtained at the time of autopsy from 3 patients who died from hypoxic respiratory failure (ages 51, 82, 97). All studies were approved by the Vanderbilt Institutional Review Board (VanderbiltIRB #'s 060165, 171657, 200490)

RNA *in situ* hybridization: RNAScope technology (ACDBio) was used to perform all RNA *in situ* hybridization (RNA ISH) experiments according to manufacturer's instructions. RNAScope probes to the following human genes were used: *TMPRSS2, ACE2, SCGB1A1, FOXJ1, SFTPC,* and *AGER*. Probes to the following mouse genes were used: *Tmprss2, Ace2, Scgb1a1, Foxj1, Sftpc, Hopx.* RNAScope probe to the SARS-CoV-2 virus (sense) was also used. Our validation of the SARS-CoV-2 probe included application of this probe to lung tissue from healthy adult lungs from a patient who died in 2018, which showed no amplification or fluorescence (Figure S8). Positive control probe (*PPIB)* and negative control probe (*DapB)* were purchased from the company and performed with each experimental replicate, with representative data shown (Figure S8).

Immunofluorescence: Using the published protocol from ACD Bio (https://acdbio.com/technical-support/user-manuals, 323100-TN), we combined immunofluorescence with RNA in situ hybridization. We used a monoclonal antibody to TMRPSS2 (Abcam, ab109131) and RNAScope probes to *TMPRSS2* and *FOXJ1* in human samples and *Tmprss2* and *Foxj1* in murine samples. Nuclei were stained with DAPI (Vector Laboratories).

Image acquisition and analysis: Fluorescent images were acquired using a Keyence BZ-X710 with BZ-X Viewer software with 40X objective. Wavelengths used for excitation included 405, 488, 561, and 647nm lines. Automated image analysis was performed with Halo software (Indica Labs); an example of cellular segmentation and determination of co-localization is demonstrated

in Figure S9. Cell area of *Tmprss2/TMPRSS2* probe was calculated as a percentage of total cell area for each epithelial subtype. Number of *Tmprss2/TMPRSS2* positive cells (defined as having 5 or more copies of *Tmprss2/TMPRSS2* by Halo analysis) as a percentage of total epithelial cells was also calculated. Presence of SARS-CoV-2 in epithelial cells was determined by Halo segmentation and analysis.

Data availability. Raw data and counts matrices used in generating this dataset are uploaded as a separate data supplement. Code used for dataset integration and analyses in this manuscript are available at https://github.com/KropskiLab.

Statistical approach: Statistical tests used for specific comparisons are described in the relevant figure legends. Comparisons of gene expression from scRNA-seq were performed by analysis of variance (one-way ANOVA) using the linear model (lm) function in R version 3.6.3. A p-value <0.05 was considered statistically significant.

References for Methods Supplement

1. Plosa EJ, Benjamin JT, Sucre JM, Gulleman PM, Gleaves LA, Han W, et al. beta1 Integrin regulates adult lung alveolar epithelial cell inflammation. *JCI Insight.* 2020;5(2).
2. Young MD, Behjati S. . SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. *bioRxiv.* 2020;303727.
3. Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36(5):411-20.
4. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. *Cell.* 2019;177(7):1888-902 e21.
5. Wolf FA, Angerer P, and Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19(1):15.
6. Polanski K, Young MD, Miao Z, Meyer KB, Teichmann SA, and Park JE. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics.* 2020;36(3):964-5.
7. McInnes L HJ, Melville J. . UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv.* 2018.

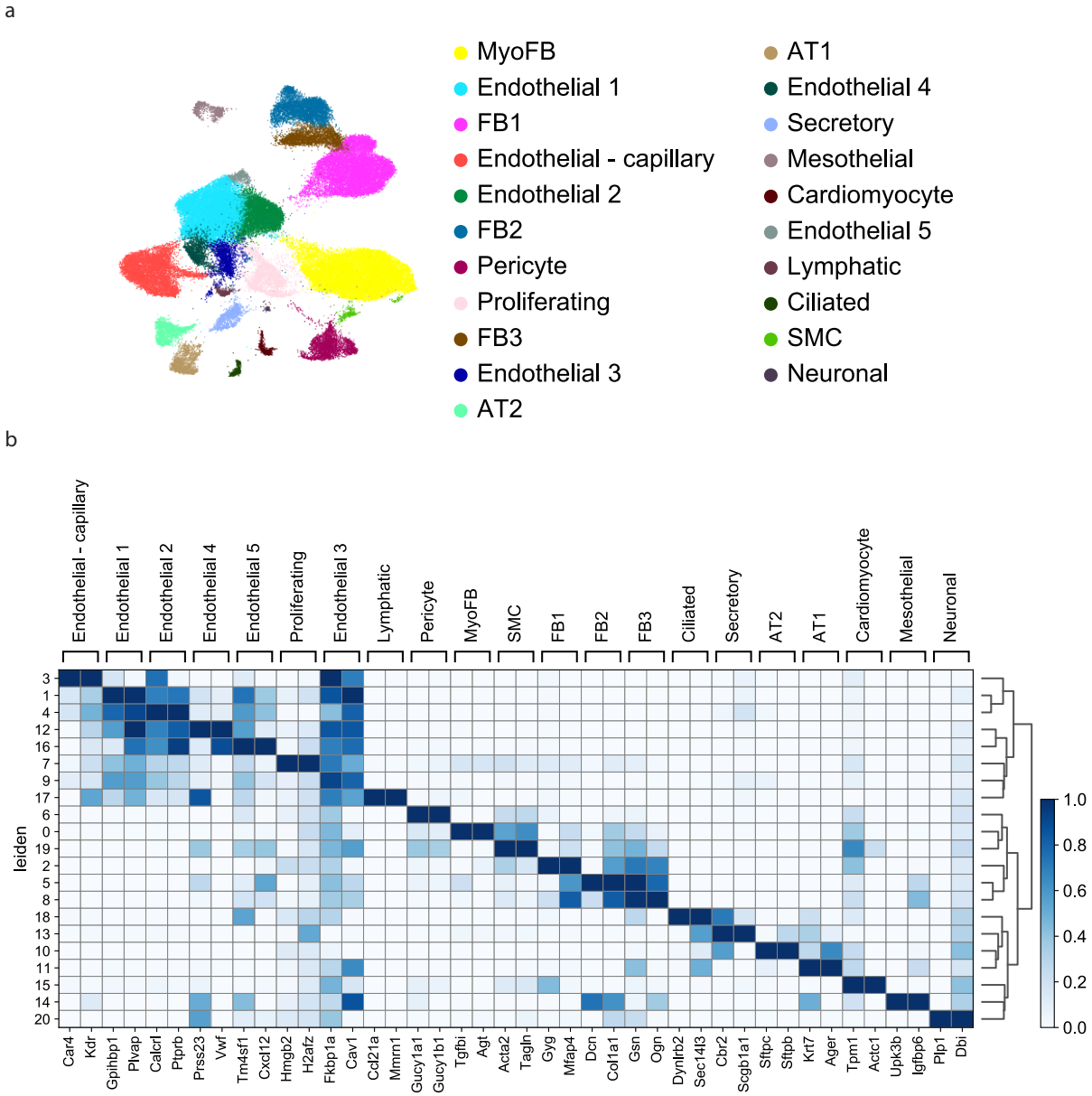

Figure S1. Cell types identified by time-series scRNA-seq. A) Enlarged UMAP embedded of 67,629 cells from across developmental time annotated by cell-type. B) Matrixplot depicting the top 2 markers for each cell-type determined by Wilcoxon-test comparing a given cluster to all others.

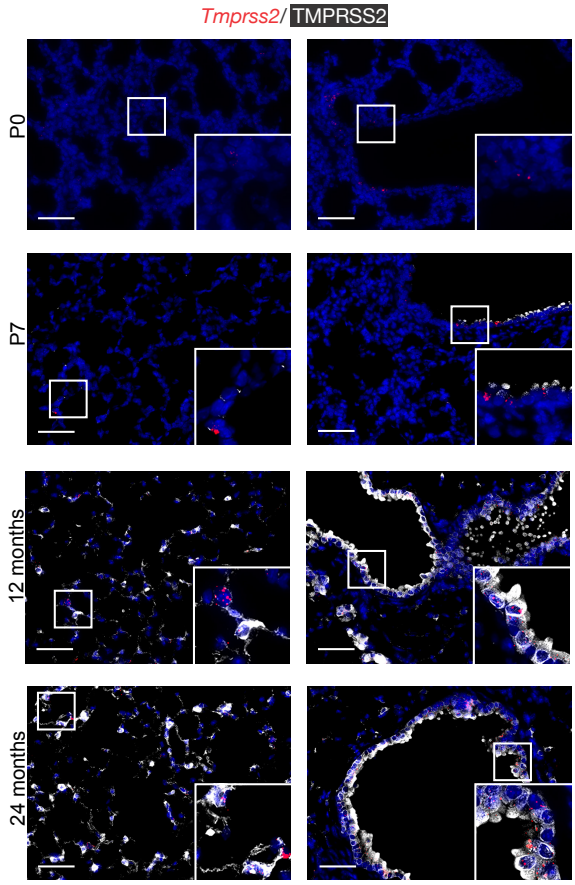

Figure S2: Validation of spatial and temporal localization of TMPRSS2 at protein and RNA level across murine lung development. RNA in situ hybridization (ISH) of *Tmprss2* expression (red) with protein immunofluorescence for TMPRSS2 protein (white). Formalin fixed paraffin embedded tissue from lungs at timepoints P0, P7, 12 months, 24 months from 3 mice at each time point were used, with ten 40X images obtained per slide. Scale bar = 100um.

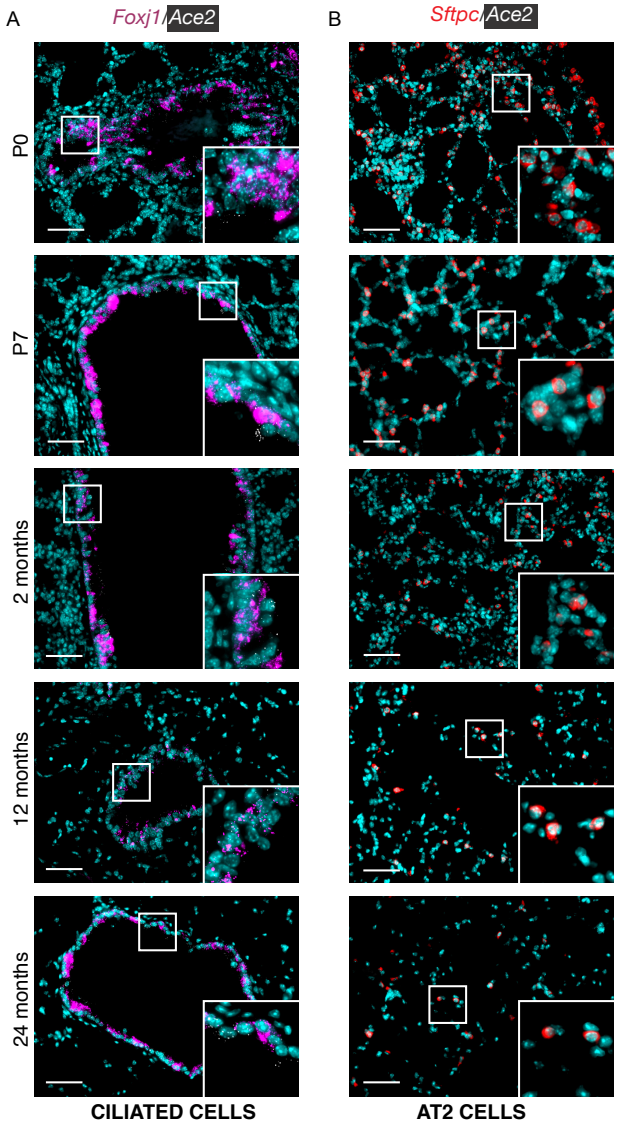

Figure S3: Spatial and temporal localization of *Ace2* expression across lung development. A) RNA in situ hybridization (ISH) of *Ace2* expression (white) with epithelial cell markers *Foxj1* (ciliated cells, magenta) and *Sftpc* (surfactant protein C, AT2 cells, red). Formalin fixed paraffin embedded tissue from lungs at timepoints P0, P3, P5, P7, P14, P64. Lungs from 3 mice at each time point were used, with five 40X images obtained per slide. Scale bar = 100um.

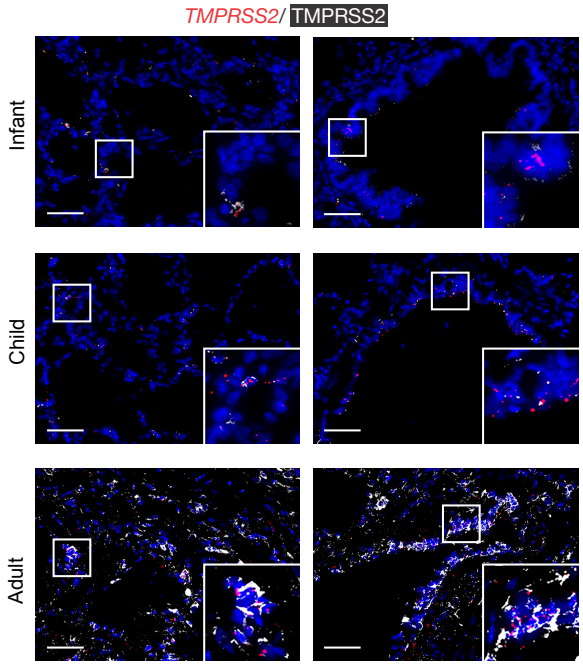

Figure S4: Validation of spatial and temporal localization of TMPRSS2 at protein and RNA level across the human lifespan. RNA in situ hybridization (ISH) of *TMPRSS2* expression (red) with protein immunofluorescence for TMPRSS2 protein (white). Formalin fixed paraffin embedded tissue from lungs between the ages of birth and 69 years was analyzed. We defined infants as birth-2 years, children 3years-17 years, and adult specimens were between 53-69 years of age. Five 40X images obtained per slide for analysis. Scale bar = 100μm.

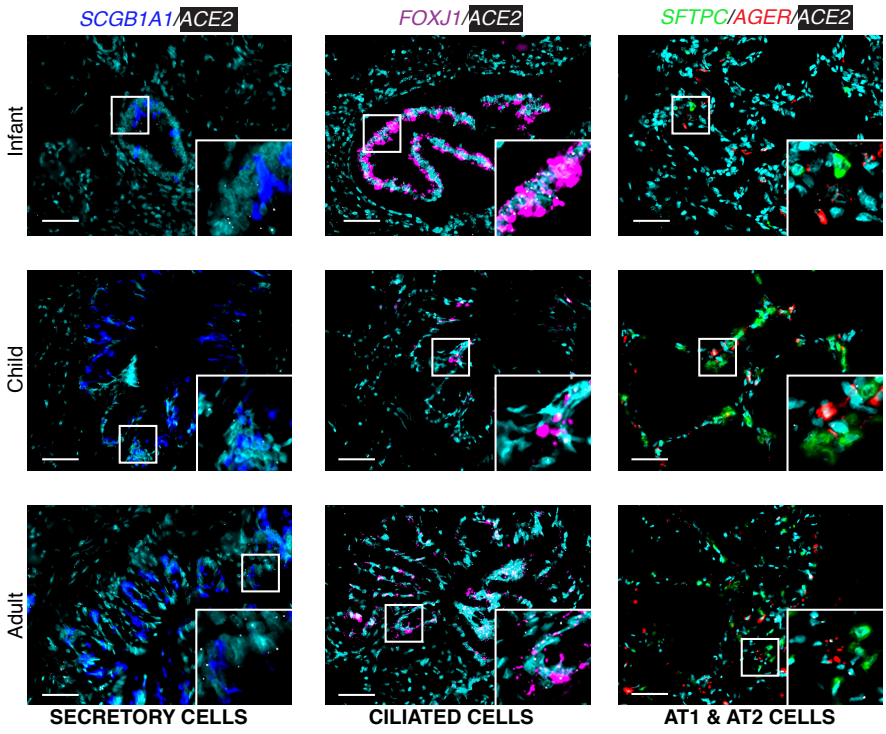

Figure S5: Spatial and temporal localization of *ACE2* expression in the lung across the human lifespan. RNA *in situ* hybridization (ISH) of *ACE2* expression (white) with epithelial cell markers *SCGB1A1* (secretory cells, blue), *FOXJ1* (ciliated cells, magenta), *SFTPC* (surfactant protein C, AT2 cells, green), *AGER* (AT1 cells, red). Formalin fixed paraffin embedded tissue from 25 human lungs between the ages of birth and 69 years was analyzed. We defined infants as birth-2 years, children 3 years-17 years, and adult specimens were between 53-69 years of age. Five 40X images obtained per slide for analysis. Scale bar = 100μm.

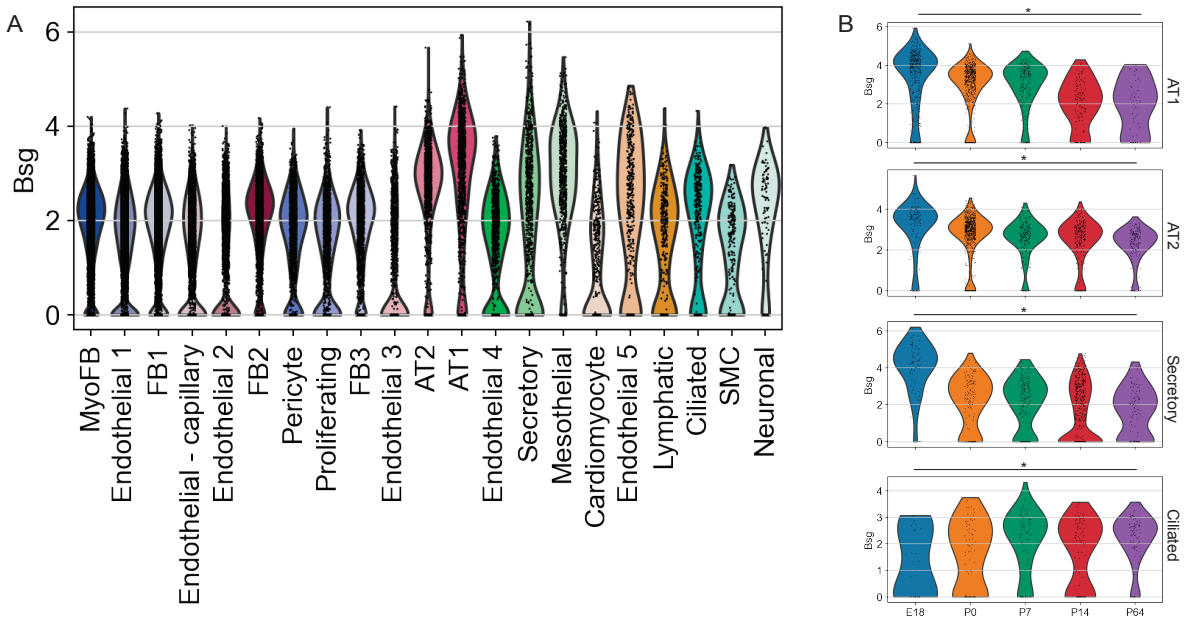

Figure S6. Expression of proposed alternative SARS-CoV-2 receptor CD147 (Bsg) across lung development. A) Violin plot depicting *Bsg* expression across cell types. B) Violin plots depicting *Bsg* expression in epithelial cell types across developmental time. * p <0.05 by ANOVA with Bonferroni correction for multiple testing.

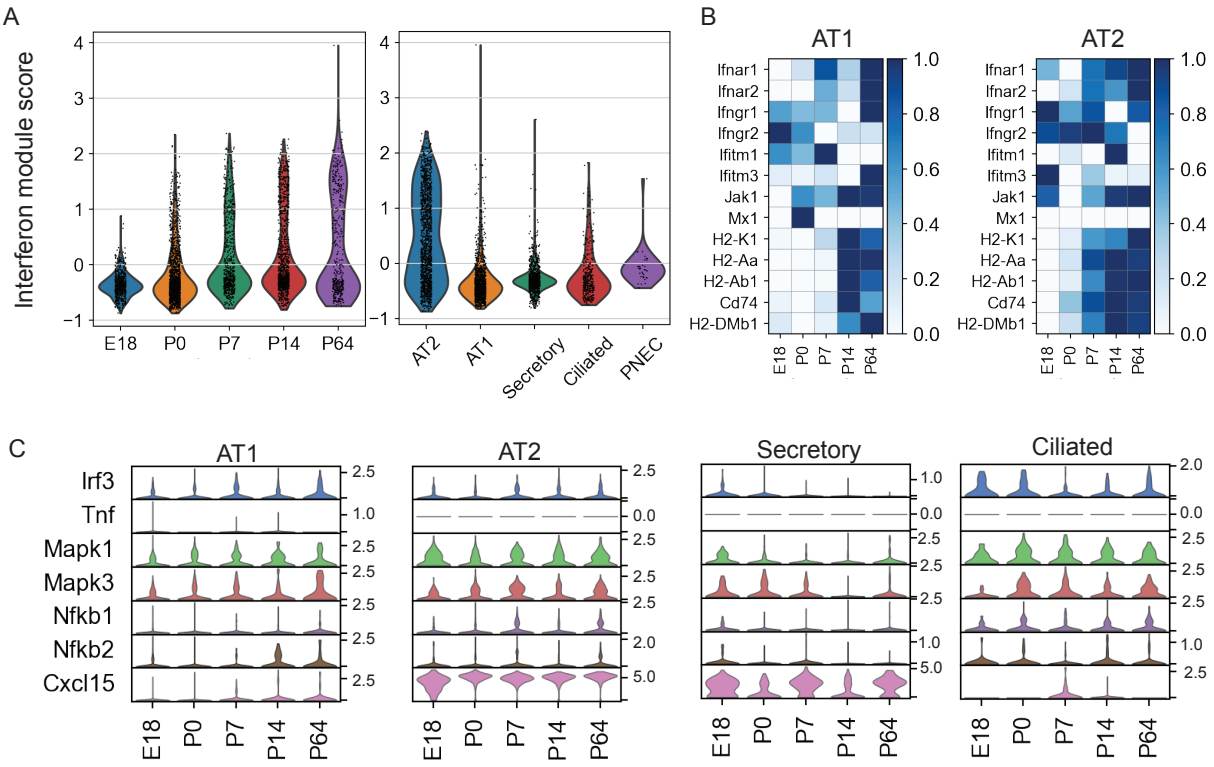

Figure S7. Immune expression programs across developmental time. Interferon gene module score was calculated using the scoregenes tool in Scanpy v1.46 using interferon receptors and canonical interferon-response genes plotted A) in all epithelial cells by developmental timepoint and by epithelial cell type. B) Heatmap depicting relative expression of interferon-related genes in AT2 and AT1 cells across developmental time. C) Expression of innate immune gene across developmental time.

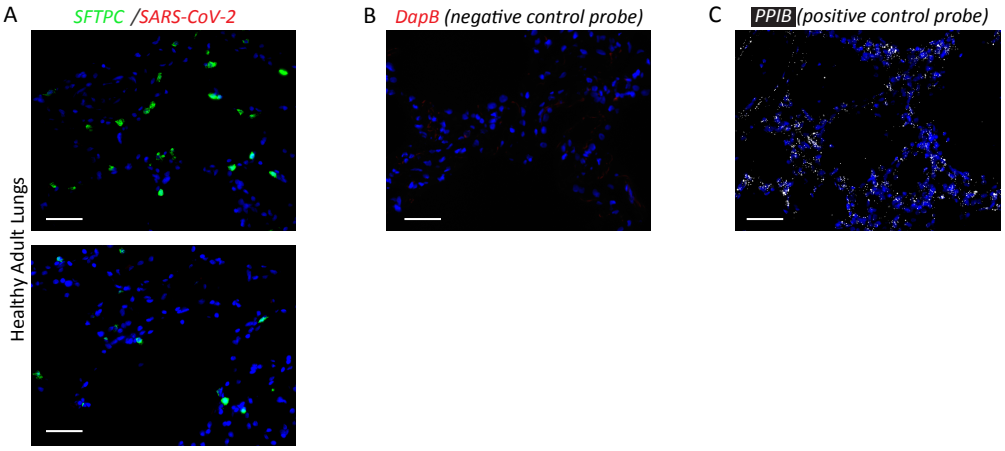

Figure S8: A) RNA *in situ* hybridization (ISH)with probes for SARS-CoV-2 RNA (red) and *SFTPC* (green) in lungs from a patient who died in 2018, which demonstrates no amplification or fluorescence of SARS-CoV-2 RNA. B) RNA ISH with *DapB* (red), a negative control probe for an *E. coli* gene provided by ACDBio. C) RNA ISH with *PPIB* (white), a positive control probe for a housekeeping gene provided by ACDBio. Scale bar = 100μm.

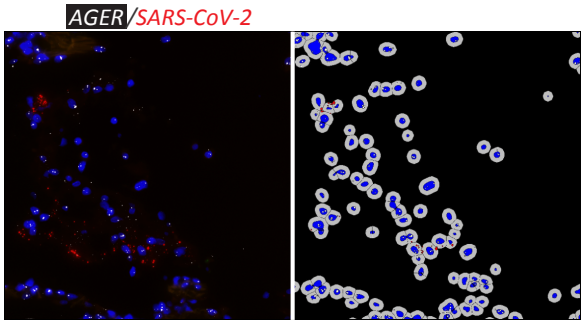

Figure S9: Example of cellular segmentation and automated labeling of epithelial cells in HALO Software as containing SARS-CoV-2 virus, shown here with *AGER* positive cells from lung parenchyma of a patient who died of COVID-19.